



Types of Search Sites

Directories for browsing.

Search Engines for pinpointing

Selecting Search Engines

- Try out several.
- Consider factors related to an engine's collecting function and its search interface.
- Use online help to get to know search features.

WWW Search Thing Authors

From the McGowan Library:

Roxanne Byrd Erin Bauer
Brenda Epperson Laura Zechin

From NN/LM-MR: Nancy Ralston

WWW Search Thing Comments:

E-mail: nralston@netserv.unmc.edu

Phone: 800-338-7657 or 402-559-4326

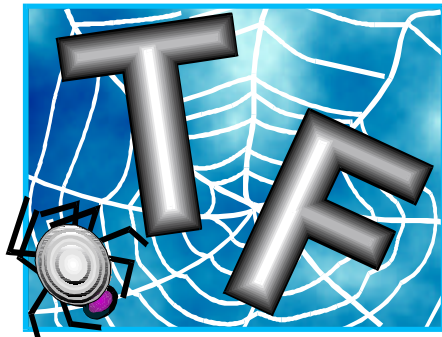


Search Site Features To Look For

SIZE	Varies significantly from one engine to another. Large engines index from 30 - 50 million web pages; small engines index around 2 million pages.
FULL-TEXT SEARCH	Engines may index every word on a web page, just an abstract, or a condensed copy of a page.
BOOLEAN	You may be able to enlarge or restrict searches using operators like OR, AND and NOT. OR increases retrieval while AND and NOT reduce retrieval. Some engines allow you to select these operators from a menu. Others allow you to enter operators along with terms, and still others insert a (default) operator between words.
PROXIMITY	Frequently, you can increase precision by specifying how words should be positioned. For example, you may be able to indicate that two words be side-by-side or within 5 or 10 words of each other. Quotes and/or the NEAR operator are often used by searchers to specify adjacency/proximity. See entry samples included in the Search Thing grid.
TRUNCATION	Truncation usually refers to retrieving several terms by entering only a word stem followed by a symbol (<i>i.e.</i> , <i>neoplas*</i> to retrieve <i>neoplasm</i> , <i>neoplastic</i> , or <i>neoplastic</i>). Truncation can be automatic, non-existent, or available only at the end of a word. Truncation symbols vary.
CASE SENSITIVITY	Case sensitive searches establish retrieval based on exact match of upper and lower case letters. Frequently, you can run case-insensitive searches by using

Search Features continued

	NO-CAPS and case-sensitive searches by including capital letters.
FIELD SEARCHING	The information about a search site that is found in a search engine's database exists as parts or fields such as the title, url, links, text, etc. Sometimes limitations may be made to these fields.
NESTING	When more than one boolean or proximity operator is used, an engine may process the entry according to a certain "batting order." Nesting allows you to group similar terms and specify the operation to be processed first. Parentheses are frequently used to indicate which part of the search should be done first.
ADVANCED FORM	Additional or different search features may be available via advanced form - only a click away.
LIMITING	Search sites such as Infoseek and Yahoo let users run one search of the entire database, then another search focusing on only documents found by a previous search.
WEIGHTED SEARCH	Some search engines allow users to control the relevance score for search results by preceding words with + or - signs to increase/decrease the importance of words entered in the search form.



Q:

True or False? A search engine finds information for its database by accepting listings sent by authors or by getting the information from automated robots sometimes called "Web crawlers" or "spiders".

A: **TRUE**

Q:

True or False? When a computerized web crawler roams the Internet looking for information to add to an engine's database, it sends back just the title and URL of each page it visits.

A:

FALSE. A crawler can send back the above, the entire text of each page, or data contained in document fields.



Search Site Components

- The **database** - where records about web pages and Internet resources reside.
- The **automated robot or indexing system** - used to gather information about sites/resources.
- The **search interface** - which allows users to tap the resources of the database (*includes search form, available search features and presentation of results*).

Presentation of Results

The way in which a search engine ranks and displays search results is not the primary focus of this document, but it is an issue well-worth examination. Many search engines return so many items that the typical user can be discouraged when it becomes necessary to sift through several results pages. Often there is an elaborate scheme established to rank items retrieved. Search engines vary regarding these algorithms, as well as the helpfulness of summaries for items retrieved.

This handout has been designed to introduce medical/health professionals and students to World Wide Web search sites and techniques. Information has been selectively chosen and will be updated twice a year by the librarians listed on page 1. To suggest changes to the publication or to request copies, contact Nancy Ralston at the National Network of Libraries of Medicine - Midcontinental Region. E-mail: nralston@netserv.unmc.edu Phone: 800-338-7657

	 OnSite Knowledge	 Your Personal Internet Guide	 just what you're looking for!							
	AltaVista http://altavista.digital.com/ Full-text. Indexes 31 million Web pages & 4 million Usenet newsgroups.	Lycos http://www.lycos.com/ Full-text. Indexes 20-25 million pages. Directory and search engine.	WebCrawler http://wc3.webcrawler.com Full-text. Indexes 2 million pages.							
FEATURES	Boolean AND (&), OR (), and NOT (!) are available (Advanced Query).. hospital and safety	Boolean AND and OR = available. AND is assumed if no operator is used. Lycos Pro menu options: <i>All the Words</i> (AND), <i>Any of the Words</i> (OR) and <i>The Boolean Expression</i> (AND, OR, NOT)	Boolean AND, OR and NOT are available. hospitals and rural hospitals or clinics							
	Proximity In Simple Query, use quotes to search phrases; in Advanced Query, use NEAR (~) to require that words be within 10 words of each other. "sleep apnea" Michael near Debakey <i>(retrieves for Michael E. Debakey, Michael E. Debakey, and Debakey, Michael.)</i>	Proximity Only available / Lycos Pro (custom search): NEAR, FOR, BEFORE and " ". NEAR requires that words be within 25 words of each other. " " searches for phrases. eye near surgery spinal near/3 injuries - Retrieves <i>spinal injuries</i> as well as <i>spinal cord injuries</i> .	Proximity NEAR, ADJ and quotes = available. ADJ and quotes specify that words must be next to each other. NEAR used with a number allows you to specify within how many words of each other the words must be. heart adj attack eye near4 surgery "carpel tunnel syndrome"							
	Nesting Use () . (drug or food) labeling	Nesting Lycos Pro: Use () . (drug or food) labeling	Nesting Use () . (drug or food) labeling							
	Truncation * retrieves word stem plus 5 characters. trial*	Truncation Automatic. A period at the end of a word turns off truncation. Use \$ to increase truncation.	Truncation Automatically strips endings. clone retrieves for <i>cloning, clones, etc.</i>							
	Field Limitations 7 are available: <table><tr><td>anchor:</td><td>applet:</td><td>link:</td><td>url:</td></tr><tr><td>host:</td><td>image:</td><td>text:</td><td>title:</td></tr></table> title:hospital	anchor:	applet:	link:	url:	host:	image:	text:	title:	Field Limitations Not available.
anchor:	applet:	link:	url:							
host:	image:	text:	title:							
FEATURES	Required/Prohibited Words Precede a word with a + to require it. Precede it with a - to exclude it. +ADA - dental	Required/Prohibited Words Use + or - before a word to require or exclude it. +ADA -dental	Required/Prohibited Words Not available.							
	Case Sensitive NO when all letters are lower case; YES when mixed case is used.	Case Sensitive NO	Case Sensitive NO, converts terms to all lower case.							
	Advanced Forms ✓	Advanced Forms ✓ Lycos Pro	Advanced Forms ✓							
NOTES	Notes <ul style="list-style-type: none">Separate help sections for Simple and Advanced search modes.Advanced Form allows limit to date.	Notes 18 WebGuide buttons (categories such as Health, Science, Technology) and Information Services buttons (i.e., PeopleFind, RoadMaps) provide browsing approach.	Notes If no Boolean operators are used, Natural Language Search is performed. This means that <i>any</i> and all words are searched, and the most relevant = returned in the first 25 listings. medical education in russia							



SITE

FEATURES

☆☆

FEATURES

☆☆

FEATURES

☆☆

FEATURES

☆☆

NOTES

InfoSeek
<http://www.infoseek.com/>

Full-text. Indexes 20-50 million pages and 15,000 Newsgroups.

Boolean AND and OR are available.
hospitals and rural hospitals or clinics

Proximity Use quotes or hyphens to search phrases.
sheep-cloning
"carpel tunnel syndrome"

Nesting Use ().
(drug or food) labeling

Truncation Automatic

Field Limitations
4 are available:

link: url:
site: title:

url:nih

Required/Prohibited Words
Use + and - symbols preceding words.

+sheep, clone -pig

Case Sensitive When caps are used to search names, search is case sensitive. If no caps are used, search is not case sensitive.

Advanced Forms ✓

Notes

- Use commas to separate lists of names.
Ian Wilmut, Roslin Institute
- Too much retrieved? Conduct 2nd search. This will be performed on initial results.
- Use | to search subtopic in one step.
sheep|clone
- Ultrasmart and Ultraseek = available.
- Ultraseek hides (Infoseek) Directory & Ultrasmart offers additional features.

Yahoo
<http://www.yahoo.com/>

Indexes web pages, email addresses, newsgroups.

Boolean - Options Page:
MATCH ON ALL WORDS (AND) and
MATCH ON ANY WORDS (OR)

Proximity - Advanced Syntax: Use " "
"mad cow disease"

Nesting
Not available

Truncation Use *.

Field Limitations
Advanced Syntax: 2 are available:

t: (title)
u: (url)

u:nih

Required/Prohibited Words
Advanced Syntax: Use + and - symbols preceding words.

+sheep, clone -pig

Case Sensitive No

Advanced Forms Advances Syntax rather than form.

Notes

- Catalog created mostly from submissions.
- To customize, go to Options Page or use Advanced Syntax.
- If search retrieves nothing, AltaVista = automatically searched.
- More than one syntax in one entry, use proper order: + - t: u: *
- Options page allows time limitations.

Sites

To Explore

Explore this site to Locate

Cybermaps **maps**
<http://www.delorme.com/cybermaps/>

City.Net **maps**
<http://www.city.net/>

Internet Grateful Med .. **MEDLINE**
<http://igm.nlm.nih.gov/> **citations & abstracts**

National Institutes of ... **NIH info.**
<http://www.nih.gov>

Internet Travel Network .. **flight info.**
<http://www.itn.net/cgi/get?itn/index>

Switchboard **people & business**
<http://www.switchboard.com/>

Lycos People Finder **people**
<http://www.lycos.com/pplfndr.html>

Lycos EmailFind **e-mail addresses**
<http://www.lycos.com/emailfind.html>

Four11 Directory **addresses**
<http://www.four11.com/>

Starting Point-New **new web sites**
<http://stpt.com/general/newsite.html>

What's New Too **web announcements**
<http://newtoo.manifest.com/WhatsNewToo/index.html>

C|NET **software & more**
<http://www.search.com/?nscp>

Alta Vista **images**
<http://altavista.digital.com/>

DejaNews **Usenet messages**
<http://www.dejanews.com/>

Handout Produced by:

Roxanne Byrd Erin Bauer
Brenda Epperson Laura Zechin
Nancy Ralston

McGoogan Library of Medicine